

Case Study: Centralised Computing Resource, University of Hull

Primary contact for case study: Vicky Mays, Records Manager and Business Systems Liaison
Information Management, Brynmor Jones Library

Additional input from: Chris Awre, Head of Information Management (Academic Services)

1. Overview

The general aim of the case study is to explore the real-world potential for implementing identified Greening Information Management (GIM) methods. Each case study undertaken will determine current information management practice across a specific information service/collection within Higher Education Institutions (HEIs). It will then assess the feasibility of implementing GIM methods within that environment and consider the costs and benefits to the organisation as a result of such implementation(s).

2. Introduction

The University of Hull was founded on 1927 (as University College Hull) and became a university in 1954. This case study is with "Academic Services" at Hull, whose purpose is to support, enable and advance the work of the University through the provision of information systems, services and resources. The focus of the case study is on centralised computing resources, specifically a shared network drive that academic services are reorganising. The network drive's information structure and use is being reorganised to facilitate internal information sharing and communication.

3. Phase 1: Examining the current IE

A large amount of information is stored on the Microsoft-based shared network drive and includes the following types:

- Project documents
- Guidance documents
- Strategic plans
- Publications
- Reports
- Policies and procedures
- Presentations
- Manuals
- Statistical documents
- Working documents

The document formats are largely Microsoft Office file types but also include PDFs and image files. This information is largely managed by individuals on an ad hoc basis. No formal information lifecycle management methodology is currently used and space is generally freed up when the drive is near-full by requests to individuals to delete files that are no longer required. Users of the shared network drive are those from the Library, Archives, IT Services and Information Management section at the University of Hull.

The Hull team has not attempted to capture the current information environment of the shared network drive by using a formal tool such as e.g. DAF (Digital Audit Framework) or DRAMBORA (Digital Repository Audit Method Based on Risk Assessment).

3.1 Stewardship requirements

Hull indicated that they retain information mostly for business purposes but much of the information will be subject to Freedom of Information legislation. They also indicated it was in their best interests not to keep information unnecessarily as this would cut down on the burden of searching for relevant information which might relate to an information request. They also noted that access to the information on the shared drive is affected by the way in which the information is stored and managed. Some documents retained on the network drive are stored for archive purposes for future reference.

There are many files stored on shared network drives which could be rationalised. Some are duplicated and many are no longer required for business purposes. At the moment these could only be identified manually and would require the help of the person who created the files to ensure that the content was no longer relevant or required. Other procedures are required in cases where those people are no longer employed by the institution or the creator of the documents is not known (orphan documents).

4. Phase 2: Evaluating techniques to green IM

Three techniques were identified as potentially relevant to Hull's current shared network drive set-up. These are pruning/weeding, centralised data storage repositories and resource limitation (quotas). The first and last of these techniques are already in place at Hull.

4.1 Manual pruning/weeding

This process would ensure that individuals evaluate the information they are keeping and that they dispose of files no longer needed. Although this is a current expectation of staff, there is no retention schedule in place to help guide the process. It is highly desirable to introduce a retention schedule to increase take-up, which is currently low. Working practices require to be changed so that the process becomes integral to daily routines. Although this process is deemed valuable, individuals seem to find it difficult to justify committing time to it, due to other activities, perceived to be of greater importance and urgency.

4.1.1 Local benefits

The main local benefit achieved through effective manual pruning/weeding within Hull's shared network drive is that the overall level of digital storage will be reduced.

In addition, the effective stewardship of local information will be increased since the awareness of individuals' understanding of what needs to be stewarded is improved. Individuals will develop a clearer idea of what they're storing and why they're storing it.

By taking control of their own information, individuals will improve the efficiency of the department. The creators of information are deemed best-placed to judge how it should best be managed. Unless information management techniques such as pruning/weeding become part of working culture, events such as staff departures will result in legacy information being left behind, which current staff do not know how best to manage. This type of scenario could be avoided by information creators regularly undertaking a manual pruning/weeding process of their own files.

4.1.2 Local disadvantages

Information could be deleted from a local shared drive, which may be required later e.g. at an institutional level. Training in information stewardship and a robust retention schedule would minimise this possibility.

4.1.3 Institutional benefits

The ability to locate relevant documents would be improved and made more efficient. Time would therefore be saved, particularly when responding to e.g. FOI requests. Legislative compliance would be improved since information should not be held beyond the time it requires to be retained for.

4.1.4 Institutional disadvantages

Pruning/weeding of information at an individual level may not be best in terms of the efficiency of the entire organisation. Different levels of granularity exist in relation to the shared drives – one at departmental level, others within that at individual level. It is likely therefore, to best protect the interests of the organisation, that different processes to implement effective pruning/weeding programmes are required, involving different people within the organisation, depending on who has the best understanding of what information requires to be retained/disposed of. It is thought that a centralised pruning model might be suited to the academic shared drive but certainly not for pruning at the level of individuals' files.

4.2 Centralised data storage repository

In some sense, the existing shared network drive does constitute a centralised data storage repository. Ownership of existing network drives is at departmental level. Each drive therefore contains copies of institution-wide documentation, thereby creating a collection of duplicate files. An enterprise-level storage repository would minimise this type of duplication and reduce the number of document versions being stored at departmental (and other) levels.

4.2.1 Local benefits

Improved centralisation would improve efficient access to, and searchability of, documents within the system.

4.2.2 Local disadvantages

Departmental staff may perceive a loss of control over their data since reliance on a central model removes the current level of control held by local departments.

A centralised repository may not be entirely suited to the effective management of all types of information currently held within the department. For example, working documents are well-suited to being held in the local network drive whilst under development. Some information, therefore, may be compromised through the total implementation of a centralised system. The centralised system would require a private, shared area with appropriate levels of access and security for accommodating departmental information (and indeed information at any level).

When accessing (or downloading) centralised files, they are often stored in a user's local temporary (temp) directory, for the purpose of opening the files locally. A policy ensuring that this directory is purged on a regular basis would minimise the impact of this local disadvantage.

4.2.3 Institutional benefits

The introduction of a centralised data repository is likely to result in the following benefits at an overall institutional level:

- Improved version control of institutional documents throughout the entire organisation.

- A reduction in the overall disk space used due to minimised duplication of documents.
- Increased legal compliance due to centralised control and location of institutional documentation. For example, FOI requests would no longer require the identification of information scattered throughout the institution.
- Reduction in staff time due to the improved efficiency in locating documents for all purposes.

4.3 Resource limitation - Quotas

Current practice shows that individuals store files on the departmental shared network drive indefinitely, for a variety of reasons including lack of incentive, lack of perceived value of removing information and fear that files may be destroyed which might be needed at a later date. Little is done to manage these files until disk space is running out. Individuals have their own quotas, which they cannot exceed. This is the only current method in place that forces individuals to review the information being stored. It is not considered an ideal method but it does drive people to review the relevancy of their files and delete what they no longer need to retain. It may be that quotas therefore have the potential to act as a means of encouraging people to undertake information management techniques such as pruning/weeding and de-duplication exercises.

4.3.1 Local benefits

Excessive amounts of redundant information cannot be stored on institutional servers. This means that individuals require to evaluate the status of information they are holding and either preserve it for the long term or dispose of it entirely.

4.3.2 Local disadvantages

Information becomes fragmented and less well-managed as individuals look for ways to get around storage problems imposed by quotas.

4.3.3 Institutional benefits

Individuals are forced to assess the information they are storing. In theory, this should make for more effective retention and disposal of files, which, in turn, should lead to increased compliance.

4.3.4 Institutional disadvantages

Due to the quotas imposed, once limits have been reached, staff often download files onto mobile devices such as flash drives, enabling them to store the information indefinitely. This poses security issues and difficulties when searching for information in response to FOI requests.

5. Phase 3: Assessing costs and benefits

The use of a centralised data storage repository was selected as the choice of GIM technique considered most appropriate to Hull's shared network drives for those working in the University Library, Archives, IT Services and Information Management.

5.1 Working practices

Should a centralised data storage repository be implemented, staff would require to be trained in the use of the new system. Although deemed intuitive since the processes largely mirror individuals' current practices, a formal training programme would incentivise individuals and make them more aware of what they are doing and why they are doing it.

Individuals may be reluctant to share their information at an enterprise level. Effective motivators and a clear vision of benefits would therefore require to be promoted throughout the institution.

The provision of secure access to different files would be essential. Assigned lists currently exist controlling who has access to different areas of the shared network drive. Varying levels of granular access to areas of the repository would be required.

5.2 Business implications

System maintenance and provision to make the system available constantly, together with system back-up, will pose implications for the institution's overall business. Increased ease of access to files, and hence compliance, will have positive business implications.

5.3 Digital footprint

The potential increase or reduction in digital footprint when comparing the current system with the requirements of a centralised data repository is unknown. It is envisaged that a centralised repository could potentially increase the institutional digital footprint, when compared with the footprint of a shared network drive. Due to the introduction of a separate system, as opposed to using a standard file system, it is possible that energy requirements will be increased.

On the other hand, enterprise level solutions could reduce the number of systems in use overall. For example, it could potentially limit the need for the existence of locally shared drives altogether (creating a reduction in hardware requirements), since a centralised solution would host a shared area capable of replacing currently used drives.

It is possible that energy consumption may be minimised within a centralised solution through the use of virtualisation. Hull already makes use of virtualisation in its institutional repository.

5.4 Change management

Change management would be central to the successful implementation of a centralised repository at enterprise level. Institutional culture is key here. The identification of selling points is required to promote change throughout the institution.

5.5 Evaluating costs/benefits

In the short-term, costs are likely to increase in relation to training, infrastructure, system support and software. In the long term, this investment will bring benefits in terms of staff costs (since staff time spent searching for information will be greatly reduced) and compliance costs. Increased functionality of the new system is likely to justify the initial cost of making the change. It remains to be seen whether these longer term benefits outweigh the initial costs incurred.

Hull want to advocate change but need to be able develop a business case and to demonstrate benefits to do so. It is difficult to do this without knowledge of current costs. There's a pressing need, therefore, to identify these in order to accurately assess the extent of benefits that might be achieved.

It is not straightforward to clearly identify where costs and benefits might be incurred. This is partly due to a large number of systems (e.g. a portal, a web content management system, a repository, a VLE) in the institution currently being used to manage content, each of which has been implemented for a very specific purpose. These should ideally be rationalised as it is now unclear where specific pieces of content should be stored. A data audit could be useful to identify potentially useful techniques for rationalising systems, introducing effective information management processes within these and for producing guidance documentation.

An information map of the institution is being drawn up. Although comprehensiveness is not being aimed for initially, the exercise will provide a broad picture of information held and is intended to further understanding of what types of information are held by different people in different locations throughout the institution. This is envisaged as a resource-level exercise, as opposed to auditing systems only. Once completed, areas for improvement should be identified and potential costs and benefits are more likely to emerge.

An interest in cost benefit analysis tools such as Insight and BILS was expressed. It is felt that a baseline is required before benefits can be clearly demonstrated and appropriate changes to working practices subsequently advocated and implemented.